House Prices Prediction

What is the framing question of your analysis, or the purpose of the model/system you plan to build?

Our field study concerns house prices in King County, USA. The county comprises houses with varied features. The features include bedrooms/house, bathrooms/bedroom, area of the house and lot, presence of a waterfront, views, condition of the house, grade assigned by the county, built year, renovated year, and the location of the house. We empirically study how the various factors influence the house prices. Our chi-square tests and T-tests analyzed the hypothesis put forward. Our regression analysis revealed the best fit model to predict the price of the house.

Who benefits from exploring this question or building this model/system?

The code within this project will assist stakeholders, i.e. home buyers or real estate agents, in predicting the sale price of King county homes while demonstrating general trends in the market.

Datasets:

For this study, we collected data from the website named "Kaggle"-(https://www.kaggle.com/harlfoxem/housesalesprediction). Kaggle is a platform for predictive modelling and analytics competitions in which statisticians and data miners compete to produce the best models for predicting and describing the datasets uploaded by companies and users.

Tools:

there are plenty of tools to help us analyzing the data and come up with a conclusion. In this project, Python language and its unique libraries such as NumPy, Pandas and Sklearn will be used to both visualize and analyze the data. Also, if time allows, Power BI will be used to create a dashboard for visualization that has a powerful graphics to make interactive visualizations. Lastly, TensorFlow is a commonly used tool for deep learning, neural networks is massive tool for modelling regression problems.